SERIEs (2012) 3:273–290
DOI 10.1007/s13209-011-0042-y

ORIGINAL ARTICLE

Anarchism, postmodernism and realism under confirmatory bias

Juan Urrutia Elejalde

Received: 5 November 2010 / Accepted: 22 January 2011 / Published online: 25 February 2011

Abstract In order to revamp Rhetoric as a methodological approach in Economics, this paper combines natural selection in evolution and the psychology of *confirmatory bias*. This latter can be thought of as a second best adaptation to the forces of natural selection and can also be an evolutionary stable strategy so that it is here to stay as seems to be supported by several psychological experiments. But once *confirmatory bias* is at work it is quite clear that economic agents in general or scientists in particular do not act as perfectly rational in the sense that they do no mimic the behavior of a Bayesian statistician. This combination has yielded three main results. First honest and open, power-free, conversations may not preclude systematic error in appreciation of theories. Therefore the moral constraint supposedly operating on the opinions of scientists might not be binding in the sense that their opinions might look completely anarchistic. Second the social constraint might also be not binding because each scientist opinion carries the same weight regardless of fame or honor, a very postmodern situation. Third, one can be a supporter of the correspondence theory of truth, one can have no doubts about the existence of an independent underlying real world and yet one might be obliged to accept that an honest and informed conversation may lead to the acceptance of false theories.

For the last 30 years or so Salvador and I have talked Economics and we have worked together in trying to build up a friendly atmosphere for its development and, more generally, for the development of science policy at different levels. The underlying current of the present paper can be understood as my last move of the chess game we are still playing around this scientific policy. I am grateful to David Teira and an anonymous referee for helping me to shape this last version of my argument.

J. Urrutia Elejalde (✉)
Fundación Urrutia Elejalde, Fortuny 37,
Sotabanco, 28010 Madrid, Spain
e-mail: juelejalde@gmail.com

Keywords Rhetoric of economics · Natural selection · Confirmatory bias

JEL Classification B41

1 Introduction

The present paper should be considered as a constructive comment *on* the Rhetoric of Economics intended to revamp it as an alternative methodological approach concerned with the acceptance and rejection of economic theories through the consideration of values and other rhetorical devices rather than objectivity. It is written from the standpoint of a theoretical economist who tries to use current theories or models in selected fields (biology and psychology) in order to clarify what this special brand of methodology is saying. The Reflexivity typical of Economics, and other Social Sciences, makes the distinction between Realism and Rhetoric rather problematic. On the one hand in my "Realismo y Economía" Urrutia (2008). I tried to really close the debate about Rhetoric and Realism opened twenty years before in *Economics and Philosophy*, confronting two papers by Uskali Mäki.[1] I argued then that, in Economics, reality could be constructed from expectations, something completely unthinkable in physics, say. The present paper shows, in addition, that Rhetoric cannot be associated with a *correspondence theory of truth*. Therefore the necessity emerges of endowing Rhetoric with some semantic force.[2]

I start by examining the debate between Uskali Mäki and Deirdre McCloskey in the *Journal of Economic Literature* (1995), somehow summarized in a later paper by Mäki (1999) and followed by a kind of addendum by him in the *Journal of Economic Issues* Mäki (2000). I want to reflect upon Mäki's diagnosis of McCloskey and his proposals for her making sense and to do so, once again, from the stand-point of an economist. But now the economics I want to rely upon is *microtheoretical* but neither completely standard (in the sense that it will be influenced by biological and psychological ideas) nor alien to the present macro discussion. As will be seen, this essay can then be taken as a second best exercise in an environment exhibiting less than perfect rationality. This combination seems appropriate for the study of Rhetoric and fits rather well the ontological constraint Mäki (2001) calls www (the "way the world works").[3] The main finding is that, in spite of the fact that there are conditions under which anarchism, postmodernism and realism are compatible among them, Rhetoric is not always compatible with a *correspondence theory of truth*.

In the next section I summarize Mäki's reconstruction and diagnosis of McCloskey (1983) as well as his proposals to make sense of McCloskey. The reconstruction

[1] I am referring here to Mäki (1988) and to Mäki (1996). In the latter he concentrated on the way Realism can be understood in Economics. In the former Mäki tried to show the compatibility of Realism in general and Rhetoric.

[2] In Urrutia (2003) I tried to explain the semantic potential of Rhetoric understood in a certain "architectural" way inspired by Sah and Stiglitz (1988).

[3] Since in these two pieces I used economic theory they can be seen not only as comments *on* the Rhetoric of Economics but also as articulating an essay *within* the Economics of Economics which can easily be extended to the Economics of Science in general or to the Economies of Art and Culture in particular.

is clear enough and the diagnosis claims that McCloskey is *not* an anarchist, *not* a postmodernist and *not* a realist. These two operations—reconstruction and diagnosis—together with the proposals that follow, will allow me to establish my purpose with greater precision. In the third section of the paper I deal with anarchism as an epistemic strategy consisting in not paying attention to any moral or social constraints in the acceptance or rejection of theories, i.e. "anything goes". In order to legitimize anarchism (or "anything goes"), I make use of a paper by Waldman (1994) on natural selection[4] in evolution to conclude that the moral constraints usually imposed on scientific conversation might have no bite, because the emergence of a systematic error makes them impossible to grasp and, hence, apply. This is a result of interest in itself and quite relevant for the discussion about the intellectual state of today's macroeconomics. In the fourth section I present some psychological experiments reported by Rabin (1998) which can provide some grounds for taking seriously *confirmatory bias* (specific kind of systematic error) as a primitive concept, and then I report on its main implication: *overconfidence*. In this next section and in the following one I make extensive use of Rabin and Schrag (1999).[5] In the fifth section I use some of their results on the implications *of confirmatory bias* (*wrongness* and *no learning*) to suggest that one can be a realist and yet entertain a coherence theory of truth instead of a correspondence theory of truth just because reality might not possibly be intellectually grasped. In the sixth section I try to legitimize postmodernism. This is a vague notion that here it is used, just like anarchism, only in its epistemic sense. Postmodernism is the indifference between the purported quality of various opinions when it comes to choosing between them. Following again Rabin and Schrag (1999), I recall some conditions under which *confirmatory bias* could lead to discount the opinion of elites and to count only on a kind of majority rule when considering problems of aggregation of experts' or scientists' information, therefore legitimizing postmodernism. This is my particular gift to Salvador since it touches on a topic becoming central to our ongoing conversation. In the last section I conclude and offer some additional comments.

2 Mäki's reconstruction and diagnosis of McCloskey

For the general purposes of the paper a summary of what is to be understood as Rhetoric seems convenient. Since McCloskey is a clear writer but not very precise it is not easy to pin down what she exactly means when pushing Rhetoric as a metatheory. Therefore a certain reconstruction is necessary. I take Mäki's reconstruction of McCloskey (see Mäki (1995)) as my starting point because McCloskey seems to accept it at least approximately in her reply (see McCloskey (1995)). Let us start then by the notion of rhetoric, being understood from the beginning that Rhetoric (with big R) is the study of rhetoric as applied to different fields. According to Mäki (1995).

[4] See also for further discussion, Dobbs and Molho (1999); Robson (2002). Both continue the discussion and offer additional references to the problem of the possible second best nature of evolutionary equilibria.

[5] Psychologists usually refere to this phenomenon as confirmation bias: for a survey see Nickerson (1998) where one can observe that psicologists call *confirmation bias* what I call here, following economists, *confirmatory bias*.

> [R]hetoric is the use of arguments to persuade one's audience in an honest conversation […]. From this perspective, rhetoric is a social process which involves (i) a persuader (speaker, writer); (ii) a persuadee or an audience (listener, reader); (iii) the aim of the persuader to persuade the persuadee; (iv) argument as the means to attain the aim (and) (v) honest conversation as the social channel of persuasion. (p. 1303).

There is little doubt that every scientific discipline and every artistic or cultural activity uses rhetoric. Therefore Rhetoric as the study of rhetoric has a broad scope and a wide range of applications. I'll focus here only on Economic Rhetoric, that is on the study of rhetoric as it is used in Economics. This Rhetoric seems to me very interesting and important for two reasons. In the first place because one way (perhaps the only way) we have to select theories is by choosing the most plausible one, that is by choosing that theory which we find most *coherent* with our present set of beliefs, and secondly (and complementarily) because our beliefs are changed by (and perhaps only by) rhetoric. The point is that *coherence* and *plausibility* are open notions quite prone to be influenced by rhetorical devices. Consequently Rhetoric is closely related to a *coherence theory of justification* which according to Mäki (1995) "suggests that all beliefs are justified by their relations to other beliefs with which they cohere". This coherence theory of justification can be seen at work these days when discussing the appropiateness of different macromodels for the understanding of the Great Recession.

Let us now turn to Mäki's reconstruction of McCloskey's theory of truth. This is indeed a delicate matter. For a *correspondence theory of truth* a statement about reality is true if it corresponds or fits with the real world. This notion of truth is clear but "inoperational" since it does not provide any procedure to find truths. A *coherence theory of truth* is, on the contrary, thoroughly "operational" because it provides in its very name a procedure to ascertain the truthfulness of a statement, that it "coheres" with other statements containing the main beliefs about the real world. The distinction does not rely upon the belief, or lack of it, about the existence of an external world, neither on the possibility or impossibility of skepticism. (Mäki, 1995, p. 1306) writes: "the truth (with small *t*) of a statement consists of its coherence with a certain set of beliefs that a privileged set of humans, obeying the cannons of *Sprachethik* end up with in an ongoing conversation". This is a *coherence theory of truth*. However McCloskey qualifies this coherence theory of truth with social and moral constraints giving rise to what Mäki calls *an elite and angel theory of truth*: only utterances coming from "good economists" matter and, among those, attention should be paid only to the ones reached in the process of an "honest" conversation.

Given this reconstruction Mäki's diagnosis follows immediacy at least in negative terms. McCloskey is *not* "an intellectual anarchist subscribing to a literal understanding of the credo 'anything goes' with respect to economics" (p. 1311) because certainly she proposes and tries to further moral and social constraints upon scientific conversation. Neither is she a *postmodernist philosopher* for whom "all participants in any conversation are on equal footing" (p. 1312), because she relies on elite's opinions. Finally McCloskey is not a "realist about truth" (p. 1312) even if she believes in the existence of the independent reality of the world.

Once reconstruction and diagnosis have been elaborated, Mäki indulges, naturally, several proposals quite attuned to his own way of looking at these matters. Mäki is a scientific realist (Mäki 1996) and therefore he dislikes a *coherence theory of truth*, but quite independently of this, and just to make sense of McCloskey's apparently arbitrary claim that Economics was in good shape (at the time of their discussion) he proposes to stick to a *correspondence theory of truth* and reserve angels and elites for the *coherence theory of justification* at most keeping them apart from any theory of truth and from Rhetoric itself. Let us look briefly at each of these proposals.

As far as Rhetoric is concerned, this proposal implies that anarchism and postmodernism cannot be excluded from Rhetoric. For a start, the absence of angels and elites does not imply, by itself, that we should dispense with Rhetoric. I will have something to add to this proposal. As for keeping angels and elites away from the theory of truth seems quite obvious since they have nothing to do with truth when one entertains a *correspondence theory of truth*. Nothing to add here. Finally Mäki proposes that at most elites and angels could be included in the *coherence theory of justification*. I will comment on that in the sequel.

The precise aim of in this paper (besides its: "celebrational" nature imposed by the occasion) can now be stated. I want to study the reconstruction, diagnosis and proposals I have just expounded in an environment (allowing for second best analysis and bounded rationality) which exhibits *confirmatory bias* and seems to fit well the undertones of Rhetoric. In this environment I want to make three broad comments. In the first place I want to reinforce Mäki's proposal of admitting anarchism and postmodernism as part of Rhetoric. However this suggestion also makes reasonable a coherence theory of truth even if one is a realist, something that Máki presumable would not like. In the second place I want to suggest that in the environment studied angels and elites are, surprisingly, compatible with anarchism and postmodernism casting doubts into Mäki's diagnosis of McCloskey. In the third place, the compatibility of anarchism, postmodernism and realism in the environment studied makes use of angels and elites in the theory of justification something more doubtful than Mäki seems to admit. I will, incidentally, also briefly express my doubts about the possibility that elites and angels could make the market for ideas to run smoothly.

3 Moral constraints (Sprachethik and Herrschaftsfreiheit)

These two notions within the parenthesis imply surely that scientific conversations are honest attempts at persuasion which exclude both lying about findings or personal characteristics and the formation of coalitions to exercise power in matters of justification. However, as we will now see these two traits might not exclude systematic errors when judging the merits of a particular theory or when trying to weigh one theory against another. The important point is that these systematic errors might look as dishonest or mafia-like behavior, these latter expressions being often heard or its content suggested in the present discussion of macroeconomic matters.

In the sequel Waldman (1994) ideas will be summarized showing that evolution might lead to an equilibrium in which the above-mentioned systematic error will appear. Let us then assume that at a point in time t of the evolutionary process there are a continuum of scientists of unit mass and a continuum of scientific institutions of unit mass. This is a technical assumption that should not bother as here. Both scientists and institutions are characterized by a pair of parameters δ and γ, $\delta \in \{\delta_1, \ldots, \delta_N\}$ $\gamma \in \{\gamma_1, \ldots, \gamma_N\}$. Now, for a scientist, $\delta \geq 0$ stands for the disutility of effort and, for an institution, it stands for the disutility of effort it inoculates in any of its members. Similarly we take γ to stand for the bias of a scientist in evaluating its own performance and for the self-confidence an institution endows any of its members with. The number and/or relevance of the contributions of scientist i is given by $F_i = f_i(e_i, I_i)$ a function of the effort s(he) chooses to make and of the quality of the institution s(he) belongs to, where quality is a continuous variable. However the subjective estimation of these contributions is given by $F_i^E = \gamma_i F_i$, where indeed γ_i is the ith. scientist bias or self-confidence. The utility of this scientist is given by $U_i = \gamma_i F_i - \delta_i e_i$. Given some standard characteristics of the function f, which make effort and quality of institution complementary, it is easy to understand that with any $\delta > 0$, the effort which maximizes utility is always smaller than the maximum possible effort $\bar{e}$.

We now turn to the exploration of both natural selection and evolutionary processes. Let us assume that at period t the data are the following. There is a proportion λ of scientists characterized by $(\hat{\delta}, \hat{\gamma})$ and a proportion $1 - \lambda$ characterized by (δ', γ').

Similarly there is a proportion of institution characterized by $(\hat{\delta}, \hat{\gamma})$ and a proportion $(1 - \lambda)$ characterized by (δ', γ'). At t there are random pairings of scientists and institutions and these pairings will determine the proportions of scientists characteristics in the next generation. Let $k(\delta, \gamma)$ be the number of "off-springs" which has a scientist with characteristics (δ, γ) under the forces of natural selection. Note that the number of off-springs depends only on the characteristics of the scientist although the characteristics these off-springs might have depend also on the institution the scientist is paired with. Let $K = \lambda k(\hat{\delta}, \hat{\gamma}) + (1 - \lambda)k(\delta', \gamma')$. Then $\hat{\lambda} = \lambda k(\hat{\delta}, \hat{\gamma})/K$ is the proportion of scientist having characteristics $(\hat{\delta}, \hat{\gamma})$ in $t + 1$. When a scientist $(\hat{\delta}, \hat{\gamma})$ "mates" with an institution (δ', γ') the inheritance yields $\hat{\lambda}/4$ scientists of each of the following pairs of characteristics: $(\hat{\delta}, \hat{\gamma})$, $(\hat{\delta}, \gamma')$, $(\delta', \hat{\gamma})$, (δ', γ').

We now turn to exploration of natural selection and the corresponding notion of equilibrium. We say that $(\hat{\delta}, \hat{\gamma})$ is a *first best adaptation* if and only if

$$k(\hat{\delta}, \hat{\gamma}) \geq k(\delta_i, \gamma_j), \quad \forall i, j.$$

Analogously $(\hat{\delta}, \hat{\gamma})$ is a *second best adaptation* if and only if

(i) $k(\hat{\delta}, \hat{\gamma}) \geq k(\delta_i, \hat{\gamma}) \quad \forall \delta_i \neq \hat{\delta}$ and
(ii) $k(\hat{\delta}, \hat{\gamma}) \geq k(\hat{\delta}, \gamma_i) \quad \forall \gamma_i \neq \hat{\gamma}$.

It is very easy to find conditions under which the first best adaptation is unique and implies $\hat{\delta} = 0$ and $\hat{\gamma} = 1$ with no discounting and no bias in the judgment of own merit. In this context the natural notion of equilibrium is *Evolutionary Stable Strategy (ESS)* which occurs when a strategy, say (δ, γ) is such that, if adopted by

all members of the population, no mutant strategy could invade the population under natural selection.

In order to formalize this notion of equilibrium in our context let $(1-\lambda)$ be the proportion of scientists and institutions with characteristics $(\hat{\delta}, \hat{\gamma})$ and let λ be the corresponding proportions with characteristics (δ', γ'). Then $(\hat{\delta}, \hat{\gamma})$ is ESS if $\exists\lambda$ small enough s. t., given any (δ', γ'), the $(1-\lambda)$ proportion of scientists characterized by $(\hat{\delta}, \hat{\gamma})$ goes to 1 as t goes to ∞. The interesting point is that $(\hat{\delta} > 0, \hat{\gamma} > 1)$ can be a second best adaptation (Waldman 1994, lemma 2) and that this second best adaptation can also be ESS provided $3k(\hat{\delta}, \hat{\gamma}) > k(\delta_i, \gamma_i)\ \forall i, j$ (Waldman 1994, Proposition 4).[6]

The very general intuition of these results (the formal proof of which is indeed in Waldman) is rather obvious. There are initial conditions in the form of population proportions in the space of characteristics which are reinforced by the forces of natural selection even when they are not the first best. As has been suggested the first best is given by a level of effort $e < \bar{e}$ and by absence of bias, $\gamma = 1$, which occurs when $\delta = 0$. As in any theory of second best if $\delta \neq 0$ the second best solution does not necessary involve $\gamma = 1$ but in general is associated with $\gamma \neq 1$.[7]

As far as Mäki's first proposal is concerned the point of the exercise is to suggest that the moral constraints of *Sprachethik* and *Herrschaftsfreiheit* might not be binding in a very particular way. In the example examined the situation is free of *Herrschaft* because the pairings are random and no coalitions are allowed to form for the mutual promotion of a particular scientific theory or approach. The situation is also such that *Sprachethik* is the rule since nobody hides its characteristics with some strategic goal in mind. However some systematic error or overvaluation of own merit might emerge through the forces of natural selection. If it does emerge we encounter a social situation which is informationally equivalent to a situation in which scientists strategically misrepresent their merits or connive with others to impose certain theories.

Under this informational equivalence is not completely foolish to entertain an anarchistic attitude. Therefore this possibility reinforces Mäki's proposal of admitting anarchism in Rhetoric, not because it is realistic to do so but because even if it wasn't the situation might look as such and there might be no way of make believe that what is going on is an honest and open discussion.

[6] Note in passing that $\hat{\gamma} > 1$ means *systematic* error. The possibility of such systematicity in committing errors contradicts the underlying logic of the rational expectation hypothesis. That is, systematic errors (in forecasting for example) can be detected and yet not to induce any revision. This seems relevant to present day discussions.

[7] The relationship between second best and evolution is interesting in itself but is also related to bounded rationality. Evolution is blind, it proceeds, for instance, by random pairings, and not by organizing agents in any optimal way. And it is precisely because it proceeds like this that evolution may fail to generate first best selection. However the solutions it does generate are persistent if they are evolutionary stable just as the QWERTY key board is persistent. But this persistence does not mean that strategies cannot change in extreme circumstances like when uniting two different populations of approximately the same size. These apparently stable strategies can also change unexpectedly when expectations have to be coordinated because there might be multiple second best solutions.

4 Confirmatory bias and overconfidence

That there might be systematic biases in the agents perceptions might be taken as a theoretical curiosity. However psychological experiments reported by Rabin (1998) seem to show quite clearly that in fact those biases are present in real life. Among these experiments the ones which seem most congenial to my purpose here are the ones referred to as *confirmatory bias*, a kind of bias which produces a certain *belief perseverance*.

Experiment 1 (Lord, Ross and Lepper).

> They asked 151 undergraduates to complete a questionnaire that included three questions on capital punishment. Later, 48 of these students were recruited to participate in another experiment. Twenty-four of them were selected because their answers to the earlier questionnaire indicated that they were " 'proponents' who favored capital punishment, believed it to have a deterrent effect, and thought most of the relevant research supported their own beliefs. Twenty-four were opponents of capital punishment, doubted its deterrent effect and thought that the relevant research supported their views". These subjects were then asked to judge the merits of randomly selected studies on the deterrent efficacy of the death penalty, and to state whether a given study (along with criticisms of that study) provided evidence for or against the deterrence hypothesis. Subjects were then asked to rate, on 16 point scales ranging from -8 to $+8$, how the studies they had read moved their attitudes toward the death penalty, and how they had changed their beliefs regarding its deterrent efficacy. At confidence levels of $p < 0.01$ or stronger, Lord, Ross, and Lepper found that proponents of the death penalty became on average more in favor of the death penalty and believed more in its deterrent efficacy, while opponents became even less in favor of the death penalty and believed even less in its deterrent efficacy. (Rabin and Schrag 1999, p. 27).

Experiment 2 (Darley and Gross).

> Seventy undergraduates were asked to assess a nine-year-old girl's academic skills in several different academic areas. Before completing this task, the students received information about the girl and her family and viewed a video tape of the girl playing in a playground. One group of subjects was given a fact sheet that described the girl's parents as college graduates who held white-collar jobs; these students viewed a video of the girl playing in what appeared to be a well-to-do suburban neighborhood. The other group of subjects was given a fact sheet that described the girl's parents as high school graduates who held blue-collar jobs; these students viewed a video of the same girl playing in what appeared to be an impoverished inner city neighborhood. Without being supplied any more information, half of each group of subjects was then asked to evaluate the girl's reading level, measured in terms of equivalent grade level. There was a small difference in the two groups' estimates—those subjects who had viewed the "inner-city" video rated the girl's skill level at an average of 3.90 (i.e., 9/10 through 3^{rd} grade) while those who had viewed the "suburban" video

> rated the girl's skill level at an average of 4.29. The remaining subjects in each group were shown a second video of the girl answering (with mixed success) a series of questions. Afterwards, they were asked to evaluate the girl's reading level. The inner-city video group rated the girl's skill level at an average of 3.71, significantly below the 3.90 estimate of the inner-city subjects who did not view the question-answer video. Meanwhile, the suburban video group rated the girl's skill level at an average of 4.67, significantly above the 4.29 estimate of the suburban subjects who did not view the second video. Even though the two groups viewed the identical question and answer video, the additional information further polarized their assessments of the girl's skill level. Darley and Gross (1983) interpret this result as evidence of confirmatory bias-subjects were influenced by the girl's background in their initial judgments, but their beliefs were evidently influenced even more strongly by the effect their initial hypotheses had on their interpretation of further evidence (Rabin 1998, pp. 27–28).

These two experiments show a psychological phenomenon called *polarization* which occurs according to the first impressions received by agents. This polarization might turn into *confirmatory bias*, a psychological trait especially well described by Lord, Ross and Lepper:

> With confirming evidence, we suspect that both lay and professional scientists rapidly reduce the complexity of the information and remember only a few well-chosen supportive impressions. With disconfirming evidence, they continue to reflect upon any information that suggests less damaging "alternative interpretations". Indeed, they may even come to regard the ambiguities and conceptual flaws in the data opposing their hypotheses as somehow suggestive of the fundamental correctness of those hypotheses (Rabin 1998, p. 28).

It should also be reported that experience and learning do not necessarily eliminate *confirmatory bias*, it may even exacerbate the problem as several experiments reported by Rabin (1998) show. This is particularly so among "experts who have rich models of the system in question". Rabin continues: "indeed many authors have hypothesized the role of reasoning process itself in exacerbating the confirmatory bias".

This *confirmatory bias* seems to be specially important under the following condition: (i) ambiguity of the evidence, (ii) abstract character of the situation, (iii) necessity of interpretation of the evidence on the situation and (iv) previous reasoning. These conditions are clearly present in the Social Sciences in general and in Economics in particular. They are all pervasive in art and culture but clearly absent in the natural sciences (or are they?) where only the fourth is applicable.

Let us now formalize the consequences of *confirmatory bias*. In this section we focus attention on how it leads to *overconfidence* (Fact 1). In the next two sections we will try to understand its eventual consequences for the pursuit of truth (however defined): Fact 2 (wrongness) and Fact 3 (no learning).

At this point we just prepare the stage for further analysis. Let us focus then on a particular agent, a scientist say, and let x be a particular theory. This theory $x \in \{A, B\}$ and $x = A$ means "x is true" and $x = B$ means "x is false". For the moment it is not

necessary to specify what kind of theory of truth we stick to. Our scientist has an *a priori* belief over the set $\{A, B\}$ which can be written as $\text{prob}(x = A) = \text{prob}(x = B) = 0.5$. At each $t, t = 1, 2, 3, \ldots$, the scientist receives a signal s_t, independent and identically distributed (i.i.d), which is correlated with the true state of the world, A or B. This signal $s_t \in \{a, b\}$. Since we do not specify whether "true" means "coherent" with other theories or "faithful" to the real world the interpretation of s_t can go from the reaction to the presentation of x to an educated audience at a seminar to the result of an experiment specially designed to ascertain the correspondence of the content of x with some trait of reality. But now, the signal s_t is not perfect and its "correctedness" can be measured by $\vartheta = \text{prob}(s_t = a/A) = \text{prob}(s_t = b/B) \in (0.5, 1)$. Note that since $\vartheta \neq 1$ we are in a context which can be called in very broad teems as one of second best. Furthermore at each t the scientist does not *perceive* s_t but rather $\sigma_t \in \{\alpha, \beta\}$ and this perception may misinterpret the signal more or less depending on the "severity" of the *confirmatory bias*. For simplicity let as assume that only those signals which conflict with *a priori* beliefs are misinterpreted and let us denote by q the probability of misreading a conflicting signal.

At each point in time the scientist updates his *a priori* beliefs according to the signals perceived. It is not very difficult to establish now a certain relationship between the "correctness" of the signal, ϑ, and the "severity" of the confirmatory bias, q, at stage t of the updating process. For so doing let $s^{t-1} = (s_1, s_2, \ldots, s_{t-1})$ and let $\sigma^{t-1} = (\sigma_1, \sigma_2, \ldots, \sigma_{t-1})$ and let us introduce two simple notions:

$$\begin{aligned}\vartheta^* &= \text{prob}\Big(\sigma_t = \alpha/\text{prob}\big(x = A/\sigma^{t-1}\big) > 0.5,\ x = B\Big)\\ &= \text{prob}\Big(\sigma_t = B/\text{prob}\big(x = B/\sigma^{t-1}\big) > 0.5,\ x = A\Big)\end{aligned}$$

i.e. the probability that the scientist misreads the evidence at t given that the previous evidence (may be also misread) has led him to entertain the wrong believe, and

$$\begin{aligned}\vartheta^{**} &= \text{prob}\Big(\sigma_t = \alpha/\text{prob}\big(x = A/\sigma^{t-1}\big) > 0.5,\ x = A\Big)\\ &= \text{prob}\Big(\sigma_t = \beta/\text{prob}\big(x = B/\sigma^{t-1}\big) > 0.5,\ x = B\Big)\end{aligned}$$

i.e. the probability that the scientist reads correctly the evidence at t given that the previous evidence (may be read correctly or may be misread) has led him to entertain the right belief. It is now very simple to establish that[8]

$$\begin{aligned}\vartheta^* &= (1 - \vartheta) + q\vartheta\\ \vartheta^{**} &= \vartheta + q(1 - \vartheta)\end{aligned}$$

[8] For example let θ^* be the probability of my misreading the signal α when B is the case but I take A to be the case because it is supported by previous evidence. It is given by the probability that the signal is correct, ϑ, times the probability that I misread it because it conflict wilt my previous belief, q, plus the probability of the signal being incorrect, $(1 - \vartheta)$ which I always believe because it agrees with my previous believes. θ^{**} can be interpreted analogously.

which are two relationships between ϑ and q. If $q = 0$ there is no confirmatory bias and $\vartheta^{**} = \vartheta$ and $\vartheta^{*} = 1 - \vartheta$. In this case the scientist has no confirmatory bias, s(he) is a perfect Bayesian statistician and, according to Harsanyi, perfectly rational. However if $q = 1$, $\vartheta^{*} = \vartheta^{**} = 1$ the scientist beliefs are completely determined by the first signal s(he) perceives, rightly or wrongly. First impression matter a lot and beliefs cannot be changed by any posterior evidence conflicting with this first impression. We could say that whenever $q > 0$ but $q \neq 1$ rationality is not perfect and this justifies my calling this paper an exercise in bounded rationality.

Let us then examine the intermediate cases. Suppose the scientist has received n_α α-signals and n_β β-signals, $n_\alpha > n_\beta$. S(he) thinks s(he) has received n_α a-signals and n_β b-signals. Therefore starting from the *a priori* belief $\text{prob}(x = A) = \text{prob}(x = B) = 0.5$ s(he) has correctly updated to

$$\text{prob}\big(x = A/n_\alpha, n_\beta\big) = \frac{\vartheta^{n_\alpha - n_\beta}}{\vartheta^{n_\alpha - n_\beta} + (1-\vartheta)^{n_\alpha - n_\beta}}$$

and to

$$\text{prob}\big(x = B/n_\alpha, n_\beta\big) = \frac{(1-\vartheta)^{n_\alpha - n_\beta}}{\vartheta^{n_\alpha - n_\beta} + (1-\vartheta)^{n_\alpha - n_\beta}}$$

such that we can define

$$\Lambda(n_\alpha, n_\beta) = \frac{\text{prob}\big(x = A/n_\alpha, n_\beta\big)}{\text{prob}\big(x = B/n_\alpha, n_\beta\big)} = \frac{\vartheta^{n_\alpha - n_\beta}}{(1-\vartheta)^{n_\alpha - n_\beta}} \lessgtr 1.$$

Λ is a kind of likelihood ratio. If $\Lambda > 1$ the scientist thinks $x = A$ is more likely than $x = B$. If $\Lambda < 1$ the scientist thinks $x = B$ is more liberty than $x = A$ in spite of the fact s(he) thinks s(he) has received more signals correlated with A than signals correlated with B. Note now that $\Lambda(n_\alpha - n_\beta)$ depends on $m = n_\alpha - n_\beta$ and that this number does not coincide with $m^{*} = n_a - n_b$. It does however when $q = 0$. In this case the perfect Bayesian statistician will calculate the likelihood ratio as $\Lambda(n_a, n_b) = \Lambda^{*}(n_\alpha - n_\beta)$. The following result can now be established

Fact 1 (Overconfidence) *If $n_\alpha > n_\beta$ and $n_\alpha + n_\beta > 1$, then*

$$\Lambda^{*}(n_\alpha, n_\beta) < \Lambda(n_\alpha, n_\beta).$$

Proof See (Rabin and Schrag, 1999, Proposition 1). □

Interpretation. Whatever is the correctednes of the signal ϑ, if there is confirmatory bias ($q > 0$) the scientist will be overconfident in his belief about which state of the world, A or B, is more likely. His or her belief in favor of $x = A$ (in favor of x being true) is stronger than is justified by the available evidence.

This is the appropriate place to soften the apparent irritation of Mäki prompted by McCloskey response in *JEL* to Mäki's diagnosis in this latter journal. In fact Mäki's

sequel in the *Journal of Economic Issues* can be read as accusing McCloskey of violating *Sprachethik*. However if we think that *confirmatory bias* is a fact we cannot infer dishonesty in McCloskey response but perhaps only some overconfidence completely coherent with *Sprachethik*.

In any case the groundwork elaborated in this section will yield its fruit in the next two, when confronting the issue of realism and elites.

5 On realism

If we want' to discuss McCloskey's claim of being a realist, and we want to discuss it in terms of truth theory we should move to a correspondence theory of truth. In this section we continue to make use of Rabin and Schrag (1999) results in order to show that *confirmatory bias* might lead to entertaining false theories at least in probabilistic terms and to eliminate possibilities of learning.

Let us first discuss the possibility of our scientist being wrong. As we already know this scientist may think that the probability of $x = A$ is $\mu > 0.5$ when in fact it is smaller than μ. S(he) is overconfident but s(he) can be said to be *right* if the true probability of $x = A$ is greater than 0.5. If in fact this true probability is less than 0.5 s(he) can be said to be *wrong*. Can s(he) be wrong in this sense? Yes if ϑ and q are sufficiently close to one.

Fact 2 (Wrongness) *There are values of* ϑ *and* q *close to one which, given*

$$n_\alpha - n_\beta \geq 2, \text{ yield } \Lambda^*(n_\alpha, n_\beta; \vartheta, q) < 1.$$

Proof See (Rabin and Schrag, 1999, Proposition 2). □

Interpretation. When *confirmatory bias* is very severe, $q \approx 1$, and the signal is very informative, $\vartheta \approx 1$, and the scientist has perceived one or none β signals s(he) is probably correct in believing $x = A$. However when s(he) has received two or more β signals s(he) is probably in correct in believing $x = A$ because s(he) does so only because the *confirmatory bias* is very severe. Therefore having received two or more β signals is something very informative while receiving additional α signals is not very informative because, with $q \approx 1$, the very first α signal is the one that explains why the scientist believes $x = A$.

Let us now turn to the possibility of learning. One might hope that, even when having received a certain amount of evidence you are still wrong according to a correspondence theory of truth, the observation of many additional signals will dissipate your ignorance and you will end up learning the truth. Rabin and Schrag have shown that this might not be the case.

Let P_w be the probability that although the true state is $x = A$, the scientist comes to believe irreversibly with near certainty that $x = B$, starting from any belief. The following fact means that there is a positive probability that this may happen.

Fact 3 (No learning) *If* $q > 1 - \frac{1}{2\vartheta}$, *then* $P_w > 0$.

Proof See (Rabin and Schrag, 1999, Proposition 4). □

Interpretation. Despite receiving an infinitive number of signals it may occur that the scientist becomes almost sure that $x = B$ when in fact $x = A$. This happens when the quantitative relationships between q and ϑ (recall $\vartheta \in (0.5, 1)$) is such that $\vartheta^* = (1 - \vartheta) + q\vartheta > 0.5$. When this is the case we already know that, once the scientist has come to believe $x = A$, he is more likely to perceive confirmatory evidence in favor of his or her incorrect beliefs than to perceive signals which conflict with this incorrect belief.

These two last facts are direct consequences of the *confirmatory bias*, a trait of the psychological personality which could very well be an evolutionary stable second best adaptation under the force of natural selection as we have seen in Section 3. Be as it may these results about wrongness and about learning problematize once more the issue of realism. Now, it is not that expectations might in fact construct reality, something I have discussed in my above-mentioned 1998 article, but rather that the correspondence theory of truth is at stake in a certain sense. One can have a realist conception of the world in the sense of believing in a world outside language (as McCloskey confesses she really does: "I'm a realist" answers to Mäki) and one can even, and correspondingly, wish to have a correspondence theory of truth (as Mäki advises McCloskey to hold) and yet face the impossibility of entertaining such a theory of truth just because the outside world may be completely inaccessible, a possibility that the two results above appear to sustain.

Under these circumstances we might easily say that "correspondence" is an interesting notion that however has to be considered inoperative. Then why should one not held a coherence theory of truth? For one thing it is accessible precisely because *confirmatory bias* has no bite against it. One might even use a more forceful argument, namely that coherence is the best strategy one may wait for in an open society so as to discover the real underlying world. This argument deserves scrutiny but it has to wait to another occasion. However some additional comments will be offered in the last section. In any case it should be clear that this argument is not immune to *confirmatory bias*.

6 On elites

In order to complete my double aim in this paper I have to show that the social constraints represented by elites might not be binding in scientific conversations. For so doing I continue to explore the consequences of *confirmatory bias* but now on how to aggregate the information provided by experts each of which judges whether $x = A$ or $x = B$, i.e. whether theory x is true or false (without restricting these notions to any correspondence with the underlying reality). Mäki notices in this respect that McCloskey puts more strength on the opinions of members of a scientific elite which in his case could be the Chicago type of economist or, more in general, the one belonging to the neoclassical tradition. However, as we will see presently, it can be shown that, under *confirmatory bias*, it is quite reasonable to pay attention to the majority of scientists or experts without weighing their opinions according to belongingness to any particular elite.

Just to be more precise, and because we are interested in Rhetoric as metatheory, let us think of a Principal (be it the metatheorist or any scientific policy board) who

must collect information regarding some particular theory from a set of scientists who may be considered his Agents. Because these agents are subject to *confirmatory bias* the optimal contract between Principal and Agents must not only take into account the usual incentive compatibility constraint but must also "prevent decisions based on good faith overconfidence" (Rabin and Schrag 1999, p. 63). In particular incentives that increase the collection of a lot of information might not be desirable.

Suppose the Principal tries to allocate a given amount of money, $W = 1$, among three different theories, T_A, T_B and T_C. Theory T_C is conventional wisdom and we can say that the "returns" of "investing" W on it are given by $r(T_C) = 1$ with complete certainty. However the other two theories are "risky" and their "returns" depend on the sate of nature $x \in \{A, B\}$ in the following way: $r(T_A/A) = r(T_B/B) \in (1, 2)$ and $r(T_A/B) = r(T_B/A) = 0$. The Principal has a conventional Von-Newman-Morgenstern utility exhibiting risk aversion and considers the possibility of asking an expert or a scientist about the most convenient allocation of $W = 1$ among the three theories. Principal and Agent share the common prior belief that $\text{prob}(x = A) = \text{prob}(x = B) = 0.5$. It is clear that if the Agent has no further information he will recommend T_C; but if s(he) gathers additional information that changes his or her *a priori* belief about the state of nature s(he) will recommend one of the risky theories.

The problem is how to incentivate the Agent. To be concrete suppose that the Agent has no cost of gathering information and that s(he) is completely risk averse. In these conditions the Principal has to pay the Agent a fixed amount and, however small, if it is positive the agent will gather an infinite number of observations. Therefore we only have two possibilities. Either the Principal pays no incentive and then, as mentioned, T_C is selected. Or the Principal pays a certain amount and the agent gather lots of information. In this latter case we now from Fact 3 that the Agent might be wrong even after gathering all the new observations. It has been shown by Rabin and Schrag (1999) that the agent, after information gathering, will identify the true sate of the world with probability:

$$\mu^*(\vartheta, q) = \frac{\vartheta\Big[2(\vartheta + q(1-\vartheta) - 1)\Big]\big(1 - \vartheta + q\vartheta\big)}{q\Big[1 - 2(1-q)\vartheta(1-\vartheta)\Big]}$$

which is increasing in ϑ and decreasing in q. Therefore the Principal's expected pay-off is

$$Eu^* = \mu^*(\vartheta, q)u(R) + \big(1 - \mu^*(\vartheta, q)\big)u(o)$$

where $u(R)$ and $u(o)$ are the utility level of income R or income zero, and s (he) does *not* want the Agent to gather information when $u(1) \geq Eu*$. Define now $\underline{\mu}$ as the value of μ^* satisfying $u(1) = \underline{\mu}u(R) + (1 - \underline{\mu})u(o)$. Since $\mu^*(\vartheta, q) \geq \underline{\mu}$ the Principal offers an incentive for the Agent to get informed if and only if $\vartheta \geq \underline{\mu}$. That is there is always a high enough q and a small enough ϑ which make not paying any incentive the optimal choice. The intuition is quite obvious, in any of these two cases the probability of being wrong is very high.

Given this intuition it should also be intuitive that if the Principal has to allocate a given amount of signals he wants to gather among several agents the more Agents the better, since each agent will be less likely to be carried away by his or her *confirmatory bias*. The next step is to think about how to weigh the information of each agent once the number of agents has been decided but the number of signals gathered by each agent is unknown. Rabin and Schrag (1999) give the following example. Suppose the correctness of the signal is given by $\vartheta = 0.6$. There are three Agents. Suppose first that they were perfect Bayesian statisticians. Two of three agents report believing in T_A with probability 0.6 (meaning that each of them has received 2 more times signals a than signals b) and one agent reports believing in T_B with probability 0.77 (meaning that he has received three more b-signals than a-signals). The Principal in this case should believe in T_B with probability 0.6. That is s(he) should pay attention to the "*strength*" of the agents beliefs perhaps weighing more heavily the believes of agents who are more experienced or belong to any kind of *elite*. What if the Principal knows that the agents are subject to *confirmatory bias*? In this case Rabin and Schrag discuss the case and come up with the following statement "If confirmatory bias is so severe that only an agent's first signal is very informative, then the Principal may wish to discount the strength of agents' beliefs and basically aggregate according to a 'majority rules criterion' " (p. 69).

It seems quite safe to say that under the conditions of this section the greater the number of scientists consulted the better and that there are no elites the opinion of which is more heavily weighted than that of non members of the elite. Scientific conversations might not be socially constrained and hence postmodernism might not be excluded from Rhetoric since all participants in the conversation might be on an equal footing. This reinforces Mäki's first proposal.[9]

7 Conclusions and final comments

This paper has articulated a second best analysis of a situation quite far from one of perfect rationality. *Confirmatory bias* can be thought of as a second best adaptation to the forces of natural selection and can also be an evolutionary stable strategy so that it is here to stay as seems to be supported by several psychological experiments. But once confirmatory bias is at work it is quite clear that economic agents in general or scientists in particular do not act as perfectly rational in the sense that they do no mimic the behavior of a Bayesian statistician. Combining second best theory and not perfect rationality—or bounded rationality—is not staple stock in Economics. However it seems appropriate to the field of Rhetoric. Quite intuitively, there does not appear to be any room for rhetoric when rationality is perfect and first best is attainable. Persuasion should be automatic and direct in such a world.

This kind of analysis has yielded three main results. First honesty and open, power-free, conversations may not preclude systematic error in appreciation of theories.

[9] In this sense it is worth mentioning that Klamer's (1983) *Conversation with Economists* is rather postmodernist in the sense that he gives equal voice to all sorts of economists. However postmodernism is a larger issue that would deserve additional distinctions between relativism and pluralism for instance. The distinction between dialogue and performance made by Mäki (2000) has also bearing on the issue.

Therefore the moral constraint supposedly operating on the opinions of scientists might not be binding in the sense that their opinions might look completely anarchistic. Second the social constraint might also be not binding because each scientist's opinion carries the same weight regardless of fame or honor, a very postmodern situation. Third, one can be a supporter of the correspondence theory of truth, one can have no doubts about the existence of an independent underlying real world and yet one might be obliged to accept that an honest and informed conversation may lead to the acceptance of false theories.

These three results obtained in a certain environment sustain the three broad comments I advanced before. First, they give Mäki additional reasons to exclude angels and elites from Rhetoric although these reasons also support a coherence theory of truth something presumably Mäki will not like. Second the compatibility of angels and elites with anarchism, postmodernism (and of course realism) can yield an alternative diagnosis of McCloskey (more anarchist and postmodernist than Mäki would concede) more attuned to her proclaimed realism. Third the compatibility of anarchism, postmodernism and realism among themselves makes room for wrongness and absence of learning and these in turn cast doubts on the necessity of maintaining angels and elites in the coherence theory of justification.[10]

These conclusions may appear as sustaining a rather severe semantic pessimism (or at least skepticism) quite attuned to the present state of Economics. With some final comments about this point I close this essay.

Semantic matters would look even gloomy if we add strategic considerations in relation to the behavior of scientists. My colleague Jesús Zamora have written on this. Zamora (1999) has described how consensus might be reached among scientists and how this consensus might be objective in the sense that scientist will not misrepresent opinions but will tell the truth. However, additional considerations of information cascades or rumors might lead to clusters of opinion completely unrelated to truth as correspondence of opinions with underlying reality.

Even if we assume away strategic considerations we have to confront the issue of whether the "market for ideas" might lead to the discovery of truth under *confirmatory bias*. The point is, however, that even if the severity of the problem were nil and even if moral constraints would underlie the market and social constraints would facilitate its functioning, the market for ideas might not work because increasing rectums to scale are very common when ideas are involved. Furthermore if we took into account strategic behavior, incentive compatibility and efficiency would require perfect competition understood as Makowski and Ostroy (2001) understand it, that is requiring full appropriation something almost incompatible with difficult appropriability of ideas. If in addition we included *confirmatory bias* it appears quite impossible to rely on the "market" as a way of reaching truth, given *wrongness* for instance.

In fact strategic behavior and *confirmatory bias* together would pose serious puzzles. Consider the following. *Confirmatory bias* might lead to every scientist to stick to its

[10] This compatibility may deprive Mäki of any consolation to his irritation about two of McCloskey's presumed inconsistencies. May be McCloskey when not really answering Mäki is not being dishonest, she might be subject to a very simple and common *confirmatory bias*. When McCloskey indulges in an unwarranted defense of neoclassical or Chicago Economics she might be misreading the evidence.

own theory and we might not discover any majority. Since majority is the only way we have to justify our beliefs under *confirmatory bias* we may want to penalize this *egoic* behavior somehow. But such an incentive scheme might generate strange dynamics specially if information flows according to rumors or cascades. (see Banerjee (1993) and Bikhchandani et al. (1992)). There are chances that clusters of false opinions might dictate what is true.

In conclusion, one cannot be very optimistic about the possibility that a *coherence theory of truth*, even including elites and angels, could be a good strategy to attain truth in a correspondence sense. The promising strategy is to study alternative imperfect social decision arrangements for their relative characteristics in relation to errors in acceptance of false theories or to errors in the rejection of true theories. What can be done is to apply to this end the literature on the architecture of imperfect economic systems initiated by Sah and Stiglitz (1988) under the possibility of *confirmatory bias*. Something I did in "La Potencia Semántica de la Retórica" Urrutia (2003).

References

Banerjee A (1993) The economics of rumours. Rev Econ Stud 60:309–327

Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as information cascades. J Polit Econ 100:992–1026

Dobbs I, Molho I (1999) Evolution and sub-optimal behaviour. J Evol Econ 9(2):187–209

Klamer A (1983) Conversations with economists. Rowman, Littlefield

Mäki U (1988) How to combine Rhetoric and Realism in the Methodology of Economics. Econ Philos 4:89–109

Mäki U (1995) Diagnosing McCloskey. J Econ Lit 33:1300–1318

Mäki U (1996) Scientific realism and some peculiarities of economics. In: Cohen RS, Hilpinen R, Qiu Renzong (eds) Realism and anti-realism in the philosophy of science. Kluwer, Dordrecht pp 425–445

Mäki U (1999) Representation repressed: two types of semantic scepticism in economics. In: Rossini Fauretti R, Sandri G, Scazzieri R (eds) Inconmensurability and translation, kuhnian perspectives on scientific communication and theory change. Edward Elgar, Cheltenhan pp 307–321

Mäki U (2000) Performance against dialogue, or answering and really answering: A participant observer's reflections on the McCloskey conversation. J Econ Issues 34:43–59

Mäki U (2001) The way the world works (www). Towards an ontology of theory of choice. In: Mäki U (ed) The economic world view. Studies in the ontology of economics. Cambridge University Press, Cambridge pp 369–389

Makowski L, Ostroy J (2001) Perfect competition and the creativity of the Market. J Econ Lit 39:479–535

McCloskey D (1983) The rhetoric of economics. J Econ Lit 21:481–557

McCloskey D (1995) Modern epistemology against analytic philosophy: a reply to Mäki. J Econ Lit 33:1319–1323

Nickerson RS (1998) Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol 2(2):175–220

Rabin M (1998) Psychology and economics. J Econ Lit 36:11–46

Rabin M, Schrag JL (1999) First impressions matter: a model of confirmatory bias. Quart J Econ 114:37–82

Robson AJ (2002) Evolution and human nature. J Econ Perspect 16(2):89–106

Sah RK, Stiglitz JE (1988) Commitees, hierarchies and polyarchies. Econ J 98:451–470

Urrutia J (2003) La Potencia Semántica de la Retórica. In: Marqués G, Ávila A, González WJ (eds) Objetividad, Realismo y Retórica. Nuevas perspectivas en metodología de la economia. Fondo de cultura económica, Madrid pp 63–86

Urrutia J (2008) Realismo y Economía. In: Perona A (ed) Contrastando a Popper. Biblioteca Nueva, Madrid pp 279–297
Waldman M (1994) Systematic errors and the theory of natural selection. Am Econ Rev 84:482–497
Zamora J (1999) The Elementary economics of scientific consensus. Theoria 14:461–488